

GREENSTONE DIGITAL LIBRARY

USING THE ORGANIZER

Dr Michel Loots and Dan Camarzan

Human Info NGO, Belgium
Simple Words, Romania

The Collection Organizer, generally called simply the "Organizer," is a freely-available software package designed to help the user create and edit the material associated with a collection. It was distributed with earlier versions of Greenstone. Its functionality has been superseded by the Librarian Interface described in the *Greenstone Digital Library User's Guide*. This document, which is provided for backwards compatibility, describes how to use it.

We want to ensure that this software works well for you. Please report any problems to *greenstone@cs.waikato.ac.nz*

Greenstone gsdl-2.50 **January 2004**



Using the Organizer

Greenstone has been used to build several collections for humanitarian purposes, collections that share a common structure and organization but with different content. The "Development Library Subset" (DLS) and "Demo" collections delivered as samples with Greenstone are examples of collections based on this structure.

The "Organizer" is an additional program distributed with earlier versions of Greenstone that assists in the creation of collections that have the DLS organization and presentation. For such collections it greatly simplified the Greenstone command-line building process. The Organizer is a Microsoft Visual C++ 6.0 application, and is therefore restricted to use under Windows (versions 95, NT or above). However, its functionality has been largely superseded by the Librarian Interface described in the *Greenstone Digital Library User's Guide*.

The Organizer is designed to help manage all aspects of organizing a digital library collection: entering document titles, assigning subjects and other metadata, altering them, etc. It works with the Greenstone digital library software by generating six intermediate files that are used by Greenstone when building collections: *collect.cfg*, *metadata.xml*, *sub.txt*, *org.txt*, *Keyword.txt* and *AZList.txt*. The structure and content of these files are explained in the *Greenstone Digital Library Developer's Guid*e.

Metadata.xml is read by the RecPlug plugin discussed at the end of Section 2.1 of the *Developer's Guid*e; it is shown in Figure 10. The *sub.txt, org.txt*, *Keyword.txt* and *AZList.txt* files define hierarchies that are used by the Hierarchy classifier described in Section 2.2; an excerpt from *sub.txt* is shown in Figure 15 of the *Greenstone Developer's Guid*e. These

five files are associated with the DLS collection (and similar collections) through the collection configuration file (*collect.cfg*): *metadata.xml* because it is read by RecPlug which is specified in the configuration file, and the other four because they are explicitly mentioned in the configuration lines that specify hierarchy classifiers.

People who want to build collections providing full bibliographic and classification functions often have difficulty in coming up with these files. The purpose of the Organizer is to help you define the contents of these files—indeed, it actually writes them for you. The Organizer is, however, limited at present to the specific document model described below, so that users who wish to add metadata not included in this model would have to make some manual modifications in the Organizer's results. A more general, flexible organizer function has been built into Greenstone's Librarian Interface; it is described in the *Greenstone Digital Library User's Guide*.

Installing and Using the Organizer

During the installation procedure from CD-ROM, early versions of Greenstone asked whether you wanted to install the Organizer at the end of the standard Greenstone installation; if you answered "yes", the installation would proceed without further intervention. If you said "no", or if you uninstalled the Organizer and later wanted to reinstall it, you should go to the *windows_utilities* directory of the Greenstone CD-ROM and run the *Organizer.exe* program.

Launch the program by selecting *Organizer* under *Greenstone Digital Library* in the *Programs* list of the Windows *Start* menu. You are asked for a database; use the DLS database by selecting the *dls.mdb* file that is supplied. You are asked for a user name and password: use *admin* for both.

Each **database** in the Organizer can contain one or more document **collections**. When you first run the newly installed Organizer, you will be presented with a list of collections that contains only one entry: the Development Library Subset (DLS) collection. Generally it is more convenient to work with the Organizer by adding new collections to the default database, including modifications or subsets of previous collections, rather than creating new databases (in fact the current version of the Organizer does support the creation of a blank database).

Each database and each collection have three basic editable **global lists**:

- *Documents* that may be included in the collection

- *Organizations*: organization names available for association with documents
- *Subjects*: subject classification terms available for building subject classification hierarchies for association with documents in a collection.

Other bibliographic attributes which can be assigned in the document model are described below.

Please note that changes to the database are automatically saved by the Organizer upon exit, without a prompt for confirmation. Thus if you wish to have a backup copy of the original content of a database, you must back it up as soon as you enter the Organizer (two backup functions are available and described below). There is no separate backup function for individual collections.

Document model

In the Organizer, each document may be assigned metadata from the following set of attributes:

i. Title
ii. Organisation (repeatable attribute generally representing the publisher, but may also be a corporate author or other organization associated with the document).
iii. Subject (hierarchical classification assignment)
iv. Keywords (additional user defined controlled indexing vocabulary —in the DLS collection it is used to classify documents according to predefined "How to" questions frequently asked by the collection users)
v. Periodical (title of the document containing the document in question, e.g. document at higher bibliographic level)
vi. Publication date
vii. Number of pages
viii. Job number (document identifier, used by Greenstone to link the metadata to the text and images of the document at the collection building stage)
ix. Language
x. Batch (identifier of a group of documents which were processed together)
xi. Suggested collections (collections for which the document is a candidate for assignment)
xii. Copyright notice
xiii. Copyright code (from a list of standard rights permission schemes)

The first five attributes are special in that the Organizer configures

Greenstone to build and display indexes based on them for the retrieval of documents. Attributes vi to viii are not indexed but are carried over as passive document identifiers in the digital library application. Attributes ix to xiii are used only for internal document management in the Organizer; they have no effect on the generated digital library application.

The subject metadata are special in that they are really a property of the collection under consideration rather than of individual documents. When a document is moved from one collection to another, all the other metadata follow, but the subject metadata must be reassigned.

It is unlikely that the developer of a digital library would wish to use the keyword metadata in the classical documentary sense, since, in addition to the subject classification, all documents are accessible in Greenstone by full text search. This attribute can be seen as simply an extra indexable attribute. It could, for example, be used for Author (not explicitly foreseen in the present model) or Country of Origin by simply placing the appropriate data in the keyword metadata. Then the button for retrieval by this new attribute (labeled "how to" in the default digital library interface) can be customised by changing the word "Howto" to "Author" or "Country" in the fifth line of the *collect.cfg* file generated by the Organizer.

Greenstone can also generate a Table of contents for access to sections and subsections of each document. This is not presently handled by the Organizer but rather must be accomplished by tagging the sections and subsections in the text itself, as described in Section 5.3 of this chapter.

It should be remembered that Greenstone itself is very flexible in terms of available document models, but that presently each different model must be formally conceived in terms of metadata functions and manually encoded in the collection building files described in section 2.2 of the *Greenstone Developer's Guide*. If the required model is relatively close to the DLS model, one could start with the build files generated by Organizer, and manually edit them according to Greenstone rules.

Exploring the Organizer

Here are details of each of the Organizer's features:

i. **The Organizer Main Window**

This window is displayed when a database has been selected. It has three parts:

- The **horizontal menu bar** at the top, including a horizontal short-cut toolbar underneath
- The **vertical toolbar** on the left
- The **central area** which displays whose content and function are selected by the vertical toolbar.

a. The **horizontal menu bar** consists of eight menus, each with one or more commands:

- The *File* menu provides only one command: *Exit* to quit the Organizer program.
- The *New* menu provides commands to add or modify Collections, Documents, Organisations, Subjects and to add or modify the lists of possible values of the other attributes in the document model. The dialogue for these commands should be self-explanatory, with the following complementary details:

  The *New/Organisation* command displays a dialogue box called *Edit Organisation Name* which provides fields to enter the full organisation name and the abbreviation (if *Auto complete* is ticked, this will be taken automatically as the first letter of each word beginning with a capital). **If the *Auto complete* function cannot propose a valid abbreviation (**because there are no capital letters in the title or because the abbreviation duplicates another in the database), **then the *Auto complete* box has to be unselected and the abbreviation entered manually.**

  The command *New/Subject* displays the *Add new subject* dialogue box to adds a category to the global (non-hierarchical) subject list for the database, whose elements are subsequently available for inclusion in the hierarchical subject classification for a given collection (*Subjects* view of the *Collection properties* window). This box can also be used to create keywords in the global subject list which can subsequently be assigned as i) the property of specific subject categories to facilitate their retrieval to add to a collection (in the *Add subject* function of the *Subjects* view of the *Collection properties* window, see below), and/or as ii) an attribute attached to specific documents (*Keywords* tab of the *Document properties* box of the *Documents* view of the *Collection properties* window, see below).

- The *Delete* menu provides commands to delete collections, documents, organisations, subjects and items from the lists of possible values of the other attributes in the document model.

- The *Edit* menu provides a *Copy* command (the same as Ctl C) which copies the headings of the central frame and the selected rows into the clipboard for pasting into files or other Organizer elements.
- The *View* menu provides commands to open the *Collections*, *Documents*, *Organisations* or *Subjects* lists in the central area (an alternate to using the top four vertical toolbar buttons). It also contains toggle boxes to display the short-cut toolbar (described below) under the menu bar and the status bar under the central frame.
- The *Tools* menu contains the following:
  - The *Documents* command contains two tools: one to globally change the name of an organisation to another in the organisation attribute of all documents in the database (the new name must first be added to the list of organisations), and another to change/add properties to a group of documents selected in the document list (the latter tool is operational if the documents concerned have already been selected in the Documents view, even if another view is presently active).
  - The second command, *Collections*, contains two tools. A *Statistics* tool to give statistical information about a given collection and an *Export lists* tool to export into a text file a list of the titles or of one of the three main attributes of a collection (the lists are saved with standard names in the *\Program Files\Human Info\Organizer\DataBase* directory).
  - The *Backup* command is used to create a back-up of the entire database each time it is selected (the database is saved under a standard name, indicating the date and time of backup, in the *\Program Files\Human Info\Organizer\DataBase* directory). Confirmation is requested so that multiple copies will not be unintentionally performed.
- The *Administration tools* menu provides windows to add new users and to delete a user or change a user's password. Only an administrator (e.g. "admin") can grant a user access to a database as "Administrator" or "Guest" and change passwords. Users and administrators have the same access rights on the collections.
- The *Help* menu contains a Microsoft standard help facility under *Help Topics* and information about the version and developers through the *About Organizer* command.

A short-cut toolbar just under the horizontal menu bar provides a set of icons which changes dynamically depending on the view which is open in the central area (set by the *View* menu—see above—or by vertical

toolbar—see below). Clicking on the second icon from the left is the same as double clicking on the item selected in the diplayed list to edit that item (see below), while the next to last icon is the same as clicking on the header of the first column of the displayed list to sort it. The first, third and last icons have the same functions as three of the menu bar commands described above: *Edit/Copy; creation of a new element in the displayed list: New/Collection/Empty, New/Document, New/Organisation* or *New/Subject; and Help/About Organizer.* Two displays associate with an additional fourth icon: permitting changing the name of the selected collection in the collections list (not in the main menu) or exporting the list of documents (same as *Tools/Collections/Export Lists* with choice of the document list, except that one can also choose the target directory and file name).

b. The **vertical toolbar** contains five buttons to select the content and function of the central area of the screen.

   The four top buttons of the toolbar are used to display for editing in the central frame a list of one of the following elements: *Collections*, *Documents*, *Organisations* or *Subjects* (these lists can also be obtained by activating the corresponding command in the *View* menu in the horizontal toolbar):

   - Collections list: Selecting the *Collections* button (the default upon entering the database) displays the list of all current collections in the database. Double click on any collection title to open the *Collection Properties* window (described below) in which you can add new documents to that collection or add/edit the collection's document attributes. To create a new collection, use the *New/Collection* command in the top menu bar or the third icon from left in the short-cut toolbar (see above).
   - Documents list: Selecting the *Documents* button displays the global list of all documents in the database (i.e. those in any of the associated collections or not yet attached to a collection). Double click on any document title to see/modify the attributes of that document other than the subject which is assigned in the Collection Properties window (described below). To create a new document entry with its attributes, use the *New/Document* command in the top menu bar or the third icon from left in the short-cut toolbar (see above).

     Text string title search: It is possible to search for titles containing a word or a string of characters by entering the word in the small dialogue box above the *Document title* heading of the list to get the first occurrence, and then clicking on the "binoculars" icon. Click on

the next button to the right ("binoculars and arrow" icon) to jump successively to the next occurrences.

Filtering Documents: Sometimes, it is convenient to view and edit only the documents of one language, one organisation or one subject. For this, select the *Filter* button ("funnel" icon) in the upper right part of the window to display the *Search documents* dialogue box, enter your preferences and accept them with the *Apply filter* button. You may also change and test the search without leaving the box with the *Search* and *Reset search* buttons. You may then apply and remove the filter at any time with *Apply filter* check box. This filter mechanism is similar to a Boolean "and" search: All the documents that are displayed will conform to all the filter "included" criteria selected.

- Organisations list: Selecting the *Organisations* button displays the global list of all organisations in the database (i.e. assigned to documents in any of the associated collections, or not yet assigned to any document). Double click on any Organisation title to open the *Edit Organisation Name* dialogue box to see/modify the Organisation name and abbreviation (same dialogue box as presented with the *New/Organisation* command). To create a new organisation entry, use the *New/Organisation* command in the top menu bar or the third icon from left in the short-cut toolbar (see above).

- Subjects list: Subjects are elements of a standard classification to be applied to the documents of a collection. Selecting the *Subjects* button enables addition, deletion and editing of the global list of all subjects in the database, even if they are not yet assigned to any document. An unlimited number of subject terms can be created for **subsequent** use at the collection level to build and assign to the collection's documents a subject classification hierarchy. Double click on any subject in the list to open the *Edit subject* window to see/modify the Subject properties. To create a new subject entry, use the *New/Subject* command in the top menu bar or the third icon from left in the short-cut toolbar (see above).

  In the *Edit subject* or *Add new subject* windows, "keywords" can be added to the global subject list (*New keyword name* box, same as using *New/Add-Modify keywords* command from the top menu bar), and can then be assigned as properties of specific subjects (*Selected for this subject* box) for use in searching for subject categories to add to a collection. The keywords (whether or not assigned to subject categories) can be used for subsequent attachment as attributes of documents ("How to" in the DLS collection). **Note that**

the two types of keywords are mixed in alphabetical order in the list which may cause inconvenience when assigning keywords to documents; this can be overcome by adding a code like "z-" in front of keywords to be assigned uniquely to subject terms (not to those to be assigned to documents unless you wish the code to appear in the generated application's keyword search list).

Warning: If the main database is very large, beware of a possible lengthy process (a minute or more) to upload collection elements. After selecting a new collection please wait until all elements are uploaded before starting to work (note that a small icon in the shape of a light bulb appears on the tab line until loading is complete). **Failure to wait for uploading may cause the program to abort.**

The *Export Files* button opens the *Export Settings* window (described below) which enables the saving of a collection's metadata to export the digital library structure into the Greenstone Digital Library software. It also enables the complete database to be saved.

ii. **The *Collection properties* window**

This window, enabling the user to build or modify a particular collection, appears when a collection has been selected in the collection list of the Organizer Main Window. It allows the user to select one of four **views** of the collection by clicking the tabs at the top. Each view in turn provides a number of data selection and editing functions which are described below:

a. *Subjects* view: An unlimited number of subject terms can be created here for the selected collection in a hierarchy up to six levels deep (although most collections will not need more than three levels). To add a subject category to the collection (to be able to use it as an attribute of one or more documents), i) select the subject heading under which you want to add the new term, ii) select the *Add subject* button, iii) select *Add Subjects from Global List*, iv) use the buttons to display the entire global list or (default and normal procedure) those which have not yet been used in the collection hierarchy, v) select the desired subject from the list, and vi) select *OK*. If the subject desired is not yet in the global list, then first use the *Add New Subject* option offered by the *Add subject* button (same as going back to the main menu and adding it with the *New/Subjects* command). Subject categories can also be modified with the *Edit* icon between the subject and document displays; the principal usage of this function is to optionally add a numbering scheme for the subject hierarchy in the categories used

in a collection (**Note that the modifications of the categories are used only for the current collection and are not carried over to the global list of subjects**).

To add one or more documents to a subject (i.e. assign the subject to one or more documents), you first select the subject category or sub-category to which you want to add documents, in the upper list box. Then select one of more documents in the lower list box, and click the little icon between the two displays with an "up" arrow and red book (or double click the documents one by one). [The documents thus assigned to the selected subject are shown with a v symbol preceding each field in the corresponding line; for convenient viewing, you can move them to the top or bottom of the display with the "v up" and "v down" icons between the displays.] You will see that the documents have been added in the appropriate place in the *Subject classification hierarchy*. Now repeat this operation until all documents are classified. One document can be assigned to as many subjects as want. You can move (but not copy) a document classified under one subject to any other subject by drag and drop with the left mouse button. To remove a document from a subject, select the document title listed under the subject, press the delete key and confirm.

b. *Organisations* view: This view is used as a convenient means of selecting or deselecting documents for the collection according to the name of the associated organisation, which will also become part of the list in the digital library for retrieval of documents by associated organisation (note that documents can also be added, and their organisation attribute changed, from the *Documents* view). The default list in the left-hand box contains the organisations already associated with documents in the collection; the right-hand box contains the titles of the documents to which that organisation has been associated—those titles marked with a v are included in the collection, those not marked are not included. Click on any title to select or deselect it in the collection. Use the icons at bottom right (white square with v and blank white square) to select or deselect all of the titles. To choose documents from organisations not yet attributed to documents in the collection, select the *Add Organisations from Global List* option of the *Add organisation* button to add organisations to the collection organisation list, then proceed to select documents as above. Similarly, you can use the *Remove organisation* button to remove all of the documents associated with a selected organisation from the collection (but not from the global document list). To work with only the documents in the collection, toggle the *List only*

selected documents checkbox. From this view, you can also add new organisations and new documents to the global list (*Add new organisation* option of the *Add Organisation* button or *Add new document* option of the *Add Document* button, the same respectively as going back to the main menu and adding them with the *New/Organisations* or *New/Subjects* commands)

c. *Documents* view: Lists all the documents that are selected for inclusion in the active collection. This is the same list of documents as the one appearing in the lower window of the *Subjects* view. The main difference is that in this document list, double clicking a given document line opens the Documents properties dialogue box for that document. You can then add/change the attributes of that document (other than the associated subject categories which are modified as described above under the *Subjects* view) by selecting one of the tabs for different classes of attributes at the top of the box:

- *General*: In this window you can enter the Document title, Job number, Number of pages and Number of images, Year of publication, and the Batch name. You can automatically input the number of images by clicking on the *Find images* button and choosing the directory containing the document's images, after specifying the possible image format extensions in the *Extensions* field.

- *Advanced*: If the document is a periodical, or if it is part of a series, you can assign the title of periodical issue or series as a property, which automatically creates an entry in the list of periodical issues which will be included in the title search window of the generated application. In the advanced window you can also assign the *Organisations* and *Languages* of the document. Both fields are repeatable: If more than one organisation is associated with the document (publishers, corporate authors) or if the document is written in more than one language, for example bilingual English/French, you should assign all of the corresponding entries from the respective lists.

- *Copyright*: It is very important to know the copyright status of a publication. This window has two parts: one where the original copyright notice is displayed and the copyright level can be assigned. This information is for document management with the Organizer, and has no effect on the Greenstone digital library application.

- *Suggested collections*: This represents a list of collections in which this document could potentially be included later. The number of

suggested collections is not limited. This information is for document management with the Organizer, and has no effect on the Greenstone digital library application.

- *Keywords*: Keywords specific for the publication can be selected. This property is an additional classification scheme which complements the subject classification and can be used in the generated digital library application to select and display documents in the digital library. In the DLS it is used for a "How to" parameter, but could also be used for any other additional metadata, for example the author or country of origin.

Note that in the *Subjects*, *Organisations* and *Documents* views, the *Add documents* button permits the user to add new documents to the collection directly from a dialogue box or by selection from the global database list. Documents added in the *Collection properties window* views are automatically entered in the global database list for future use. When a new document is added to the collection from the global database list, the *Search documents* dialogue box appears to enable the user to easily locate the desired documents according to several selection criteria (this is the same filtering function described above in discussing the *documents* list obtained by selecting the *Documents* button of the vertical toolbar of the Organizer Main Window).

d. *Others* view: The *Other hierarchies* view shows the ordering of documents under periodical titles (which cannot be edited) and also the alphabetical classification of titles under each language. The user can change the alphabetical classification groups (example A-C, E-G or A-L, M-Z etc) to the best size to display the documents in the finished library. For this, click on a language and use the *Split letters* button.When you are pleased with the result, click on the *Save Splitters* button (until then you can get back the initial or previously saved letter categories by clicking Load/Refresh Splittings or eliminate the letter categories by clicking *Eliminate Splittings*).

iii. **The *Export Settings* window**

This window, which is presented when the bottom icon of the vertical toolbar of the *Organizer Main Window* is selected, allows you to export the result of your work, and is normally the last stage involved in creating a new collection or sub-collection. Select *Export Files* to invoke the *Export Settings* window and select a collection for export and a folder to receive the exported information. Then, click *Export file*s. This will place the five files *collect.cfg*, *metadata.xml*, *sub.txt,*

org.txt, Keywords.txt and *AZList.txt* into the selected folder. In order to build the collection with this information, you need to move the files to the appropriate place. Place the *metadata.xml* file in the collection's *import* directory and the others in the collection's *etc* directory.

Getting Started in Ten Steps and 15 Minutes

The best exercise to get the look and feel of the Organizer is by actually creating a small test library. If you have 15 minutes please follow these ten steps and you will understand this program much better.

Before getting started (assuming you want to use the Organizer output to build your first Greenstone digital library), you should have:

a. Installed Greenstone (see the *Greenstone Installer's Guide*) which includes the Demo library in DLS format and its source files. **Note, if you wish to be able to add to your collection any of the 140 documents in the DLS collection in the default Organizer database (instead of just the 11 of these documents in the Greenstone Demo collection), you should install DLS as one of the sample Greenstone libraries and replace "demo" by "dls" in the instructions below.** The Demo and DLS collections will be installed respectively in *c:\program files\gsdl\collect\demo and c:\program files\gsdl\collect\dls*. If you previously installed Greenstone without DLS and wish to install it, then you may uninstall and reinstall Greenstone specifying this collection.

b. Created the structure for your new collection (we will refer to it below as "newcol") by executing the following command from *Run...* in the Windows *Start* menu:

   "c:\program files\gsdl\bin\windows\build" newcol

c. Replaced the default *collect.cfg* file created by the previous step with that used by the Demo collection. That is, copy *c:\program files\gsdl\collect\demo\etc\collect.cfg* to *c:\program files\gsdl\collect\newcol\etc\collect.cfg*. This is necessary because Demo (and all DLS format collections) use some specialised options that the default collection does not (see the *Greenstone Developer's Guide* for further details).

We suggest that you print the instructions below and follow them step by step:

1. Open the Collection Organizer, select the dls database with and give "admin" for both the user name and password (the *Collections* button of the vertical toolbar should be highlighted by default; if it is not, then select it).

2. Select the *New/Collection/Empty* command in the horizontal menu bar on the top of the Organizer Main Window to create a new empty collection. Give it the title and the version number you choose, for example "My First Collection" as the name and "1.0" as the version.

3. For certain document attributes, you will have to create a list of the possible values first. So if you know the languages and/or publishing organisations of your documents, use the commands *New/Add-Modify languages* and *New/Organisation* to add all the languages you will use in this or future collections, as well as the publishers of your documents. You can use the same commands to add or modify languages and organisations when you want, but not when you are editing the collection itself as in the next instruction.

4. Double Click on the line with the title of the collection you just created.

5. Click on the *Subjects* tab at top (if not already selected as it should be by default); then click on the *Add subject* button and the command *Add new subject*, then enter a few subjects in the *Subject title* field, pressing the "enter" key after each one. Click on the + before the word *Subjects* in the Subject hierarchy list to see the subjects you have entered.

6. Click on the *Documents* tab; this will open the Documents view. Then add documents to your collection as follows:

   a. To add a document from the Demo collection (or the DLS collection if it is installed in Greenstone) to your new collection, click the *Add documents* button and select *Add document from global list*. Locate the document you require (using the "filtering" function described above) and add it to your collection in the Organizer. After adding the document locate its source files in the Demo import directory (*c:\program files\gsdl\collect\demo\import*) and copy them to the import directory of your new collection. For example, to add the document "Butterfly Farming in Papua New Guinea" to your new collection you should take note of the document's job number once you locate it in the Organizer. This document's job number is "b22bue", so you should copy the

"b22bue" directory from *c:\program files\gsdl\collect\dls\import\ac01ne* to *c:\program files\gsdl\collect\newcol\import\ac01ne*.

b. To add a new document (that is, one that is not included in the Demo collection) to your new collection click the *Add documents* button and select *Add new document*. Enter the document's title, job number (your choice), number of pages, publishing organisation, languages and other information. You must then create a new directory in *c:\program files\gsdl\collect\newcol\import* to correspond to the job number of the new document. Into this new directory you should place the source file of the document and any associated image files (in HTML or any other format accepted by Greenstone (see *Greenstone User's Manual*.)

7. Go back to the *Subjects* tab, and you will see that your documents are displayed in the lower list box. Select a document. Then select in the upper list a subject you want this document to be classified under and click the little icon between the two displays with an "up" arrow and red book. Once a document is thus classified, you can still move it from one subject to another by drag and drop with the left mouse button. You can also move documents or subjects up or down within the same level of hierarchy by selecting the blue up or down buttons to the right of the *Subject hierarchy* list. Try to classify an average of 6 to 30 documents per subject. One document can be assigned to as many subjects as want.

8. Repeat the previous steps by adding new subjects, and adding more documents. When the library is finished, you will have to review the complete list of subjects and documents, to be sure that all are entered and correctly classified and ordered.

9. Finally, close the *Collection properties* window and press the *Export Files* button of the vertical toolbar. This will open the *Export Settings* window. Click on the *Display collection list* button, and select your collection, then click on the *Browse for folder* button and select the directory to which you want to export the metadata files, and press *Export files* to export the collection metadata for collection building with Greenstone.

10. Copy the exported files to the appropriate places in your new collection's directory structure.

    a. The exported *metadata.xml* file should be copied to the *c:\program files\gsdl\collect\newcol\import directory*.

b. The exported *AZList.txt*, *Keyword.txt*, *sub.txt*, and *org.txt* files should be copied to the *c:\program files\gsdl\collect\newcol\etc* directory.

Note that the collect.cfg file generated by the Organizer is not required because the "classify" lines it contains are already included in the collect.cfg files for the DLS and demo collections.

The newcol collection is now ready for building. Build it from the command line using the import.pl and buildcol.pl commands (see the *Greenstone Developer's Guide* for details).